It's about time: cesarean sections and neonatal health

Costa-Ramón, Ana María ; Rodríguez-González, Ana ; Serra-Burriel, Miquel ; Campillo-Artero, Carlos

Abstract: Cesarean sections have been associated in the literature with poorer newborn health, particularly with a higher incidence of respiratory morbidity. Most studies suffer, however, from potential omitted variable bias, as they are based on simple comparisons of mothers who give birth vaginally and those who give birth by cesarean section. We try to overcome this limitation and provide credible causal evidence by using variation in the probability of having a c-section that is arguably unrelated to maternal and fetal characteristics: variation by time of day. Previous literature documents that, while nature distributes births and associated problems uniformly, time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections. Using a sample of public hospitals in Spain, we show that the rate of c-sections is higher during the early hours of the night compared to the rest of the day, while mothers giving birth at the different times are similar in observable characteristics. This exogenous variation provides us with a new instrument for type of birth: time of delivery. Our results suggest that non-medically indicated c-sections have a negative and significant impact on newborn health, as measured by Apgar scores, but that the effect is not severe enough to translate into more extreme outcomes.

DOI: https://doi.org/10.1016/j.jhealeco.2018.03.004

Originally published at:
Costa-Ramón, Ana María; Rodríguez-González, Ana; Serra-Burriel, Miquel; Campillo-Artero, Carlos (2018). It's about time: cesarean sections and neonatal health. Journal of Health Economics, 59:46-59.
DOI: https://doi.org/10.1016/j.jhealeco.2018.03.004

It's About Time: Cesarean Sections and Neonatal Health[☆]

Ana María Costa-Ramón[a,b,*], Ana Rodríguez-González[a], Miquel Serra-Burriel[c,b], Carlos Campillo-Artero[d,b]

[a] *Universitat Pompeu Fabra*
[b] *Centre de Recerca en Economia i Salut (CRES)*
[c] *Universitat de Barcelona*
[d] *Servei de Salut de les Illes Balears*

Abstract

Cesarean sections have been associated in the literature with poorer newborn health, particularly with a higher incidence of respiratory morbidity. Most studies suffer, however, from potential omitted variable bias, as they are based on simple comparisons of mothers who give birth vaginally and those who give birth by cesarean section. We try to overcome this limitation and provide credible causal evidence by using variation in the probability of having a c-section that is arguably unrelated to maternal and fetal characteristics: variation by time of day. Previous literature documents that, while nature distributes births and associated problems uniformly, time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections. Using a sample of public hospitals in Spain, we show that the rate of c-sections is higher during the early hours of the night compared to the rest of the day, while mothers giving birth at the different times are similar in observable characteristics. This exogenous variation provides us with a new instrument for type of birth: time of delivery. Our results suggest that non-medically indicated c-sections have a negative and significant impact on newborn health, as measured by Apgar scores, but that the effect is not severe enough to translate into more extreme outcomes.

Keywords: Cesarean section neonatal health time variation instrumental variables

1. Introduction

Recent years have seen increasing concern over the rise in cesarean section births. Among OECD countries in 2013, on average more than 1 out of 4 births involved a c-section, compared to 1 out of 5 in 2000 (OECD, 2013). This rise has been largely debated because c-sections are associated with greater complications and higher maternal and infant mortality and morbidity compared to vaginal births. However, the available studies may suffer from omitted variable bias, as mothers who give birth by c-sections may be different from those who have vaginal births in terms of characteristics that can affect the health outcomes of the child and the mother after birth. Along these lines, the WHO has recently pointed out the need for more research in order to better understand the health effects of cesarean sections on immediate and future outcomes, remarking that "the effects of cesarean section rates on other outcomes, such as maternal and neonatal morbidity, pediatric outcomes and psychological or social well-being, are still unclear" (WHO, 2015).

This paper aims to help fill this research gap by providing new evidence of a causal link between unplanned cesarean sections and newborn health outcomes. Understanding the impact of c-sections on neonatal health is of relevance, as fetal and neonatal outcomes have been shown to be determinants not only of future health, but also of other later life outcomes, such as test scores, educational attainment, and income (Almond and Currie, 2011). In particular, we look at the impact of c-sections on Apgar scores, a widely used measure of newborn well-being. Apgar scores have been found to be predictive of health, cognitive ability, and behavioral problems of children at age three (Almond et al., 2005), of reading and math test scores in grades 3-8 (Figlio et al., 2014), and of school attainment and social assistance receipt after age 18 (Oreopoulos et al., 2008). We also analyze the effect of c-sections on other indicators of newborn wellbeing, such as needing reanimation or being admitted to the intensive care unit (ICU).

In order to show the existence of a causal relationship between unscheduled c-sections and health, we use exogenous variation in the probability of having a c-section at different times of day. Indeed, although nature distributes births and associated problems uniformly, some studies have demonstrated that time-dependent variables related to physicians' demand for leisure are significant predictors of unplanned c-sections (Brown, 1996). Using a sample of birth registries in public hospitals in Spain, we first document that, in this context, unplanned c-sections are more likely to be performed in the early

[☆] We are grateful to Libertad González, Guillem López-Casasnovas, Cristina Bellés-Obrero, Andrés Calvo, Rosa Ferrer, Christian Fons-Rosen, Borja García Lorenzo, Albrecht Glitz, Sergi Jiménez-Martín, Gianmarco León, Vicente Ortún, Alexandrina Stoyanova, Alessandro Tarozzi and Ana Tur-Prats. We also thank participants in the UPF LPD Seminar, VI EvaluAES Workshop, 31st ESPE Conference and 12th iHEA Congress.

[*] Corresponding author: Ana María Costa-Ramón. Department of Economics, Universitat Pompeu Fabra. C/ Ramon Trias Fargas 25-27, 08005 Barcelona (Spain).

Email address: anamaria.costa@upf.edu (Ana María Costa-Ramón)

hours of the night (from 11 pm to 4 am). We discuss how the structure of medical shifts and the higher opportunity cost in terms of time that vaginal deliveries imply might explain physicians' incentives to perform more c-sections during this time of day. We then show that mothers giving birth at different times of day are observationally similar, also in terms of pregnancy and labor characteristics that might predict a medically-indicated c-section. The results thus suggest that the excess number of c-sections observed at the early night are due to non-medical reasons. We consequently adopt an instrumental variable approach, using time of birth as an instrument for the mode of delivery. In other words, we estimate the local average treatment effect of c-sections on neonatal health for mothers whose mode of delivery is affected by time of birth. This allows us to interpret our estimates as causal and to focus on avoidable c-sections, as medically-indicated cesareans will be performed independently of the time of birth. Our results suggest that these non-medically indicated c-sections lead to a significant worsening of Apgar scores of approximately one standard deviation, but we do not find effects on more extreme outcomes such as needing reanimation, being admitted to the ICU or on neonatal death.

In order for our instrument to be valid, it must satisfy two conditions: first, that there is no selection of mothers with different characteristics giving birth at different times of day and, second, that giving birth during the early hours of the night only affects infant health through the increased probability of having a c-section. The comparison of maternal and pregnancy characteristics across times of day provides reassuring evidence regarding the first assumption. In order to support the validity of the exclusion restriction and, in particular, to show that variation in quality of care across time is not driving our results, we perform a robustness check restricting the analysis to births that take place during the night. Moreover, section 5 includes further supplementary tests that support our interpretation of the findings.

This paper contributes to two different strands of the literature. First, we contribute to studies on the effects of c-sections on newborn health outcomes. A large number of papers have documented a robust association between c-sections and respiratory morbidity, both at birth (Zanardo et al., 2004; Hansen et al., 2008) and in the longer-term in the form of asthma (Davidson et al., 2010; Sevelsted et al., 2015).

To the best of our knowledge, the only paper that endeavors to identify the causal impact of cesareans on later infant health is Jachetta (2015)[1]. The author uses variation in medical malpractice premia at the Metropolitan Statistical Area (MSA) level in the US as an instrument for the rate of risk-adjusted cesarean sections and finds that higher rates lead to an increase in the rate of total hospitalizations and of hospitalizations that present asthma. Although the author identifies several potential threats to the validity of the instrument, the paper is a first step towards providing evidence of the causal link between c-sections and health outcomes. We advance the existing knowledge by using a new instrument that allows us to credibly isolate the causal impact of non-medically indicated c-sections on newborn health. In particular, our setting allows us to focus on mothers that give birth in the same hospital and have similar observable characteristics, differing only in the time of delivery. Moreover, because we measure the impact on health at birth, we are able to establish a direct connection between c-sections and health outcomes.

Second, our work is also related to the literature that documents or uses time variation in the probability of having a c-section. Brown (1996) was one of the first to show that the probability of unplanned c-sections is non-uniformly distributed across time. Using data from military hospitals in the US, the author finds that cesarean sections were less likely to occur during the weekend and more likely from 6 pm to 12 am. He interprets these results as evidence that non-clinical variables, in particular physicians' demand for leisure, also play a role in doctors' decision-making. In our setting, we find that the probability of unplanned c-sections is higher during the early hours of the night. It is during this time that doctors appear to have a higher incentive to perform a c-section when facing ambiguous cases, as the opportunity cost in terms of time for a vaginal delivery is higher.

There is one paper that uses time variation in the probability of having a c-section to study maternal outcomes. Halla et al. (2016) use administrative data from Austria to show that the probability of a c-section birth is lower on weekends and public holidays. They use this as an instrument for mode of delivery, and find that c-sections reduce subsequent fertility and that this translates into an increase in maternal labor supply over a period of about six years. Our paper also makes use of time variation but our data allow us to use finer variation and rule out potential exogeneity problems: we study mothers in the same hospital, on the same day, but giving birth at different times. Moreover, we are also able to precisely identify and restrict our sample to non-scheduled c-sections.

The structure of the rest of the paper is as follows. In the next section we provide background information on the choice of mode of delivery, on the institutional setting and physicians' shifts, and on why we would expect to find an adverse effect of c-sections on health outcomes. The third section introduces the data, describes the variation in the c-section rate across a 24-hour cycle and presents the empirical strategy. In section 4 we show and discuss our results. Section 5 presents some robustness checks and supplementary analysis and, finally, section 6 concludes.

2. Background

2.1. Choice of the mode of delivery

Cesarean sections can be performed for several reasons and at different lengths of pregnancy. First, c-sections can be scheduled in advance – also known as planned c-sections – when

[1]Recent work by Jensen and Wüst (2015) and Mühlrad (2017) examines the impact of medically necessary c-sections on health for a particular group of at-risk babies: those in breech position at term. Their findings suggest positive short and long-run effects of medically indicated cesareans for this group.

there are medical indications that make a vaginal delivery inadvisable. Examples of such indications include multiple pregnancies with non-cephalic presentation of the first twin or placenta previa (NICE, 2016). In principle, c-sections can also be scheduled if they are demand-determined; that is, if the mother requests to deliver via a c-section. However, in the context of public hospitals in Spain, these elective c-sections are very uncommon and are not, in fact, included in the portfolio of services offered by the public system (Marcos, 2008). In any case, we exclude scheduled c-sections from our sample as these women are likely to be different from those delivering vaginally.

If there is no scheduled c-section, an attempt of vaginal delivery begins with the onset of labor or medical induction. If an immediate threat to the life of the woman or fetus emerges, a c-section should be performed as quickly as possible (NICE, 2011). However, some indications such as dystocia (failure to progress or cephalopelvic disproportion) have a more imprecise diagnosis which leaves the door open to a more discretionary interpretation and present large variability among clinicians (Fraser et al., 1987; Barber et al., 2011). Therefore, in some cases, whether or not a c-section is needed is not obvious, and the choice between a vaginal delivery or a c-section will depend on the subjective assessment of the doctor. Unfortunately, our data does not contain the specific indication registered by the medical team to justify the c-section. However, given that emergencies should be uniformly distributed across time, we expect any observed time variation in the c-section rate to be due to indications falling in this gray area.

As Shurtz (2013) points out, a c-section is a common procedure known to be sensitive to physician incentives. Several papers have found, for example, that financial fees can influence doctors' behavior (Grant, 2009). When fees are higher for a c-section than for a vaginal delivery, physicians have a greater incentive to perform a c-section. Other studies suggest that physicians perform more c-sections as a defensive strategy reflecting a fear of malpractice lawsuits (Baicker et al., 2006; Currie and MacLeod, 2008; Jachetta, 2015). Finally, physicians have more incentives to perform c-sections when the opportunity cost of time is higher, as vaginal deliveries take longer than c-sections and thus the latter can be seen as a time-saving device (Lefèvre, 2014). We focus here on this last type of incentive given that, by performing our analysis within hospital and exploiting variation across time of day, we abstract from variations in malpractice premia and financial fees.

In particular, the average duration of vaginal deliveries among first-time mothers is around 11 hours (NICE, 2014). The first stage of established labor[2] usually lasts about 8 hours and is rarely longer than 18 hours. After that, birth is expected to take place within 3 hours of the start of the active second stage[3]. In contrast, a c-section takes much shorter; in general the average duration of this procedure is between 30 and 75 minutes (NICE, 2014). The baby is usually delivered in the first 5-15 minutes, with the remaining time being used for closing the incision (APA, 2017). Moreover, complications during this procedure are very uncommon. According to NICE (2011), c-sections increase the risk of hysterectomy (14 more per 100,000) and of cardiac arrest (15 more per 10,000). Therefore, given the low risk in terms of complications and the expected time gain, doctors may have larger incentives to perform a cesarean section when the opportunity cost of time is higher.

2.2. Mechanisms: the impact of c-sections on newborn health

Cesarean sections have been associated with several adverse health outcomes for newborns. Hyde et al. (2012) provide an extensive review of such findings, concluding that although further research is needed, the available evidence suggests that "normal vaginal delivery is an important programming event with life-long health consequences." More specifically, the absence or modification of a vaginal delivery has been linked to several health alterations, which they classify as either short- or long-term. In what follows we summarize some of these findings, in particular those that are more relevant to understand how c-sections might affect our outcome variables. Before doing so, however, it should be noted that any negative health effect of c-sections is outweighed by its benefits when there is a clear medical necessity. For instance, in the case of breech babies, Jensen and Wüst (2015) find that c-sections decrease the probability of having low Apgar scores and the number of doctor visits in the first year of life. More generally, cesareans save lives when severe complications arise during birth.

The adverse short-term outcomes with which c-sections have been associated include the increased risk of impaired lung functioning and altered behavioral responses to stress. With regard to the former, one of the most common causes of respiratory distress among newborns is transient tachypnea or the presence of retained lung fluid. While in the amniotic sac, a baby's lungs are filled with amniotic fluid, but during labor the baby releases chemicals which, together with the pressure of the birth canal on the baby's chest, help expel the amniotic fluid from their lungs. This process does not occur when babies are born by cesarean section, such that the presence of fluid in their lungs after birth is more common. Moreover, catecholamines, one of the chemicals released by the fetus during labor, are also correlated with muscle tone and excitability. Otamiri et al. (1991) find that babies born by cesarean section responded worse to neurological tests a few days after birth. In our setting, we can proxy the impact of c-sections on these outcomes by looking at Apgar scores at minute 1 and 5 after birth, which capture, among other aspects, respiration, reflexes and muscle tone. Severe effects, in particular serious respiratory morbidity, could also be reflected in increased need for assisted ventilation or ICU admission (Grivell and Dodd, 2011).

In the longer-term, cesarean births have also been associated with a higher risk of asthma (Sevelsted et al., 2015). While one possible mechanism is change in infant microbiome as a result of not passing through the birth canal, Hyde et al. (2012) also highlight that altered lung functioning at birth may lead to

[2]Mothers are considered to be in the first stage of established labor when the cervix has dilated to about 4 cm (NICE, 2014).

[3]The mother is considered to be in active second stage of labor when either the baby is visible, or the full dilatation of the cervix has been accomplished and one of the following conditions is satisfied: either the mother has expulsive contractions or there is active maternal effort.

the development of future respiratory problems. Finally, there is evidence that the reduction in excitability among cesarean newborns may be a sympton of further alterations in the programming of the central nervous system, as affected by the catecholamine surge at birth (Boksa and Zhang, 2008). These findings generally suggest that any health worsening at birth we detect may have long-lasting consequences.

2.3. Institutional setting

2.3.1. Childbirth in Spanish public hospitals

In Spain, maternity care coverage is universal under the provision of the Spanish National Health Service. Antenatal and postnatal care for women are mainly provided at local health centers by midwives, while deliveries are supervised in hospitals by teams of both midwives and obstetricians. Expectant women do not have a pre-assigned doctor or midwife for the delivery. Rather, they are assigned to the professional available at the time of admission to the hospital. During labor, women are assisted by midwives who monitor the baby, check how labor is progressing, and call a doctor if they notice any issues. If no complications arise, midwives might manage the whole delivery. However, the obstetrician is in charge of any instrumented assistance and makes decisions regarding the mode of delivery.

Women may opt for private care, but most deliveries – 8 out of 10 births – take place under the public health system (Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015). Pregnant women are in general assigned to give birth at the hospital that is closest to their residence. In big cities where there are several public hospitals, mothers can request a change in the assigned hospital through an administrative procedure. However, hospitals in our sample are located in medium-size towns in which there are no other public hospitals.

In the year 2014, the c-section rate in the public health system was 22.1%, lower than the 25.4% rate of the whole sector, combining both public and private hospitals (*ibid.*). It is important to note that within the public system, obstetricians' wages are independent of the method of delivery used or the number of c-sections performed.

2.3.2. Physicians' shifts

In our setting, the typical work shift for a doctor is from 8 am to 3 pm; night shifts are covered by doctors that are on duty and must stay in the hospital for 24 hours (from 8 am to 8 am next morning). All doctors younger than 55 are required by law to work these longer shifts (Ministerio de Trabajo y Asuntos Sociales, 1997). When doctors are on duty, they provide assistance in (relatively uncommon) gynecological emergencies, occasionally monitor mothers' health after birth, and are present in the labor room when decisions regarding a delivery are made, or if complications arise. Midwives, on the other hand, work 12-hour shifts (from 8 am to 8 pm).

For all of the hospitals in our sample, there are at least two obstetricians and two midwives on duty during the night, and each doctor assists on average between 1 and 2 deliveries per night. During these times, each delivery thus accounts for a major part of a doctor's duties. Although in our setting doctors cannot leave the hospital while they are on duty, beds are available to rest when there is no emergency or complication that requires their presence (Ministerio de Sanidad y Política Social, 2009).

3. Data and methods

3.1. Description of the data

Our data consists of all 6,163 birth records from four public hospitals in different Autonomous Regions in Spain during the years 2014-2016[4]. The characteristics of the hospitals in our sample are comparable to that of the majority of public hospitals in Spain, in particular with regard to the volume of births attended per year (between 300 and 1500). In terms of c-section rates, three of the four hospitals are in the left tail of the distribution, while one is just at the mode, with a c-section rate around 21%. This comparison can be found in figure A1 in the appendix.

Each birth registry contains information on the mother's characteristics (age, nationality, education, marital status, etc.), on the pregnancy, on the type of birth (planned cesarean, unscheduled cesarean, eutocic delivery, etc.), on medical interventions during labor, on a series of medical indicators collected before, during, and after the delivery, on the newborn (birth weight, Apgar scores, etc.), and on the date and time of birth. Table A1 shows some summary statistics of the variables of interest[5]. In our data, 5% of women delivered via a planned c-section, more than 11% via an unplanned c-section, and 68% had an eutocic delivery, that is, a vaginal delivery without other interventions (i.e. spatula, forceps, or vacuum). Vaginal deliveries with such interventions represent around 15% of the sample. We eliminate non-single births, planned c-sections and breech vaginal babies[6]: our final sample consists of 5,783 observations.

Our main outcome variables are Apgar scores at minutes 1 and 5 after birth. These result from the examination of the health status of the newborn performed by the midwife or the pediatrician one and five minutes after birth, respectively (AEPED, 2014)[7]. In particular, they assess and grade between

[4] Data collection was approved and financed by the Spanish Ministry of Health under the Strategy for Assistance at Normal Childbirth in the National Health System (PI/01445).

[5]For comparison, in table A2 we show descriptive statistics of the coincident variables reported in the Spanish National Statistics Institute birth registries for all births that took place in hospitals in Spain in the years 2014-2015. We see a slightly higher proportion of non-Spanish women in our data and also less multiple pregnancies, but similar characteristics in terms of age, gestational length or birth weight.

[6]Breech vaginal babies – that is, babies that were in breech position and were born vaginally – are a rare case: we only have 8 of those in our sample. This is because attending such type of birth requires special caution and expertise (American College of Obstetricians and Gynecologists, 2006) – most fetus in breech position are delivered by planned c-section. Therefore, these kind of births are not a plausible counterfactual for unplanned cesareans.

[7] In general, Apgar scores can be determined by a pediatrician, a midwife or a nurse present in the labor room – this depends mainly on the routines of each hospital. In the hospitals in our sample, this task is normally assigned to midwives.

0 and 2 points each of the following aspects: appearance (skin color), pulse (heart rate), grimace (reflex irritability), activity (muscle tone), and respiration. These variables thus take values between 0 and 10. We study both the levels of these scores and also the probability of the scores being below different thresholds. We also look at whether the newborn needed reanimation (assisted ventilation), whether they were admitted to the intensive care unit, and at the event of neonatal death.

Some other medical variables included in our analysis need further clarification. Besides the outcome variables presented above, another one of interest is the umbilical cord pH, which is an indicator of fetal distress. A sample of blood from the umbilical cord artery is collected after cord clamping, and the levels of pH are measured. There is some variation in the literature in what is considered the range of normal values for this outcome, with thresholds for acidemia (low pH) spanning from 7 to 7.20 (Malin et al., 2010). In our analysis we consider thresholds of 7.20, 7.15, and 7.10. A related variable is the fetal scalp pH or intrapartum pH, which is a measure of fetal distress during labor, before birth. In this case, the pH is measured from a sample collected from the baby's head when it becomes visible. Too low values of this variable – in particular, pH lower than 7.20 – suggest that the baby is not getting enough oxygen, and thus a cesarean section might be necessary (SEGO, 2005). Finally, one relevant control we include in our preferred specifications is obstetric risk. This is recorded by the medical professionals who prepared our data, and defined as a dummy variable that takes value one if, during pregnancy, some risk factors were detected that could lead to an adverse pregnancy outcome[8].

3.2. Variation in the c-section rate by time of day

Figure 1 shows the c-section rate at different times of day for our sample of public hospitals in Spain. We can observe that the distribution of unscheduled c-sections by time of birth is not uniform. The proportion of women that deliver via an unplanned c-section is higher in the early hours of the night (from 11 pm to 4 am), and much lower during the remaining hours of the night and the rest of the day. This pattern is not matched by either the total number of births or the number of vaginal births (see figure A2 in the appendix). More importantly, this variation is not driven by differences in maternal or pregnancy characteristics of the deliveries that take place at different times of day. In the next section, Table 1 confirms the balance of a very large set of mother and pregnancy characteristics between women delivering in the early hours of the night and during the rest of the day. As we will discuss in further detail, this allows us to use this exogenous variation as an instrument for mode of delivery.

We are not the first to document this early night spike in unscheduled c-section deliveries. For example, Fraser et al. (1987), Brown (1996), and Spetz et al. (2001) show an increase in the probability of a c-section at the end of the day up until midnight, and Hueston et al. (1996) documents a peak in the unplanned c-section rate between 9 pm and 3 am. These authors have interpreted these evening or night peaks as evidence that convenience and doctors' demand for leisure influence the timing and mode of delivery. Similarly, several studies find that the probability of a c-section increases when doctors can go to sleep or return home after the birth, likely linked to the fact that cesarean sections require on average less total time devoted to the patient (Klasko et al., 1995; Spong et al., 2012).

This explanation is consistent with the time pattern that we observe in our data. Given the medical shift structure and the larger time-cost of surveillance implied by vaginal deliveries, doctors' incentives to perform c-sections in ambiguous cases may vary by time of day. In particular, we expect doctors to have a larger incentive to perform c-sections in the early hours of the night. By this time, on-duty doctors have already been working for more than 12 straight hours (see Figure A3 in the appendix[9]). If they perform a c-section and do not have other mothers to care for, they can expect to rest for the remainder of their shift. Alternatively, if they do not perform a c-section, they will need to occasionally monitor the vaginal delivery throughout the night. Moreover, ongoing deliveries in the early hours of the night have a high probability of falling under the responsibility of the doctor on duty[10], as opposed to deliveries which begin later and are more likely to finish past the doctor's shift. These conditions would suggest that a higher share of deliveries with ambiguous indications end up as cesarean sections during the early hours of the night, as compared to the rest of the day. Consistent with this interpretation, we find that the probability of doctors performing a c-section at these times increases when there is only one ongoing delivery at the beginning of the night, that is, when the expected marginal gain of a c-section is larger[11].

Other alternative explanations are not compatible with this variation. For example, if either patient's or physician's fatigue increased the probability of c-sections, we would expect to see a higher unplanned c-section rate during the late hours rather than the early hours of the night. We can also rule out that this is driven by an accumulation of births during these hours,

[8] More specifically, obstetric risk was defined as the presence during pregnancy of one or more of the following factors that increase the chance of an adverse pregnancy outcome: cholestasis, chorioamnionitis, 486 diabetes insulin and non-insulin dependent, chronologically prolonged pregnancy, multiple pregnancy, hellp syndrome, hypertension, isoimmunization in pregnancy, stained amniotic fluid, fetal malformation, uterine malformation, fetal malposition, myomectomy, oligoamnios, previous preterm labor, placenta praevia, plyhydramnios, preeclampsia, premature rupture of membranes, siphylis, toxoplasmosis, previous c-section, repeated abortions, previous miscarriages, anteparturm alteration of fetal wellbeing.

[9] Figure A3 shows the proportion of unplanned c-sections as a function of the number of hours worked by physicians: 0 hours corresponds to 8 am. As can be seen, the proportion of c-sections starts to increase when doctors have been working for already 12 hours, and reaches its maximum when hours worked are between 15 and 20. The proportion of unplanned c-section decreases in the last hours of their shift.

[10] Average duration for the first stage of labor in vaginal deliveries among first-time mothers is around 8 hours (NICE, 2014), and for the second stage around 3 hours. See section 2.1 for more detail.

[11] Table A3 in the appendix shows that the increase in the probability of cesarean birth at the early hours of the night (from 11 pm to 4 am) is larger in days when there is only one birth at night compared to days with more than one birth.

as we do not observe the same time pattern for the number of births (see figure A2 in the appendix). Finally, the early night spike in c-sections cannot be explained by selection of highly interventionist doctors at different times of day, as deliveries are not pre-assigned to a given obstetrician. We also provide evidence that this is not the case in Figure A4 in the appendix[12], where we show that there are no systematic differences among doctors in the probability of attending births during the early hours of the night.

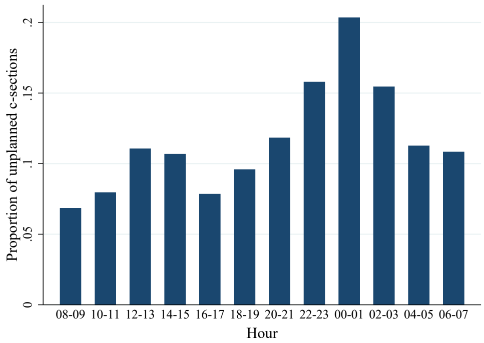


Figure 1: Proportion of Unplanned C-sections by Time of Day

Notes: The figure represents the proportion of unplanned c-sections by time of day over the sample of unplanned c-sections and vaginal births. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

3.3. Identification strategy

Our objective is to identify the causal impact of non-medically indicated c-sections on infants' health at birth. The simple comparison of women who had a c-section and those who delivered vaginally is likely to suffer from omitted variable bias, as these groups likely differ in characteristics that influence the outcome variables. Table A4 in the appendix compares observable characteristics of these two types of mothers. We observe, in fact, that these mothers are significantly different in terms of several relevant aspects such as age, gestational length, obstetric risk, or educational achievement, all potentially related to the health of the newborn. There are thus reasons to be concerned that they might also differ in other characteristics we cannot observe. Moreover, a comparison of vaginal deliveries and births by c-section does not allow to identify which kind of c-section is causing whatever health effects are found, since we observe the outcomes of both medically and non-medically indicated interventions. In order to overcome these issues, we use variation in the probability of having a c-section by time of day. The purpose of the instrument is thus twofold: to compare similar women, and to precisely identify the impact of non-medically indicated cesareans.

We define a binary variable CS_i equal to one if the mode of delivery is an unplanned c-section and zero if it is a vaginal delivery (eutocic or operative). Infant health H_i refers to either Apgar scores or other measures of neonatal health. We would thus like to estimate the following equation:

$$H_i = \beta_0 + \beta_1 CS_i + \beta_2 X_i + \epsilon_i \tag{1}$$

where X_i is a set of covariates that include information on mothers' personal and pregnancy characteristics. As discussed earlier, the estimation of equation (1) is, however, likely to provide biased estimates of β_1. To overcome this potential endogeneity, we use an IV approach, instrumenting the type of birth with an indicator for the time of day the infant is born. Therefore, our first stage is as follows:

$$CS_i = \gamma_0 + \gamma_1 earlynight_i + \gamma_2 X_i + \upsilon_i \tag{2}$$

where $earlynight_i$ is an indicator variable equal to 1 if woman i gives birth during the beginning of the night (from 11 pm to 4 am). We expect a positive $\hat{\gamma}_1$ since obstetricians are more likely to initiate a c-section during these hours of the night in order to gain time for rest or leisure.

The identifying assumption is that $earlynight_i$ is not correlated with ϵ_i, but this assumption entails two conditions. The first is that the instrument is as good as randomly assigned. We provide suggestive evidence that this is the case by comparing personal and pregnancy characteristics of mothers who give birth between 11 pm and 4 am and those during the rest of the day in Table 1. Mothers are similar with respect to their age, educational level, weight and height, alcohol and tobacco consumption habits during pregnancy, gestational length, obstetric risk, weight of the newborn, or previous c-sections. The level of intrapartum pH, a measure of fetal distress during labor – a major cause of emergency c-sections – is also equivalent. Mothers are also comparable in terms of the average time that they have been in the hospital, that is, time between admission and time of birth. We find some slight differences between mothers across time of day with respect to nationality (there are slightly more non-Spanish women during the day shift) and marital status (more unmarried women during the day). However, these differences are very small in magnitude. We also find that the proportion of women whose labor was induced is higher during the early hours of the night (28.5%) compared to the rest of the day (22.6%). This is something one might expect from our institutional setting, since in the hospitals in our sample most inductions are performed in the morning and, given the average duration of labor, these women are more likely to give birth during the early hours of the night. We control in our main specification for all of these differences and perform a robustness check excluding inductions in Section 5.2, where we find that our conclusions still hold. Overall, we thus feel confident with the assumption that there is no selection of women into the different times that could threaten our identification.

Additionally, identification requires the exclusion restriction to hold; that is, the instrument should affect infant health only through the increased probability of having a c-section. One potential concern is that the quality of medical care could change

[12] Figure A4 plots, for a small sample of births for which we know the doctor who attended the delivery, the probability of attending births during the early hours of the night across different doctors.

Table 1: Maternal Characteristics by Time at Delivery

	Means		p-value
	Rest of the day	Early night	for difference
A. Personal characteristics			
Mother's age	31.729	31.888	0.349
Level of education			
No school	0.033	0.025	0.146
Primary school	0.254	0.262	0.563
Secondary school	0.525	0.523	0.906
University education	0.187	0.189	0.876
Non-Spanish	0.256	0.223	0.015
Single	0.019	0.009	0.017
Mother's weight	65.561	65.779	0.630
Mother's height	1.650	1.607	0.534
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.126	0.606
Alcohol during pregnancy	0.004	0.004	0.891
Gestation weeks	39.263	39.274	0.853
Previous c-section	0.090	0.103	0.173
Obstetric Risk	0.388	0.409	0.161
Intrapartum pH*	7.271	7.278	0.402
Birth weight	3277.356	3270.303	0.662
Induction	0.226	0.285	0.000
Time in hospital (in hours)*	9.891	10.156	0.450
Observations	4478	1305	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies). Variables marked with an asterisk (*) are not available for the whole sample. Intrapartum pH is only available for a sample of births (425 observations), and time in hospital is only available for one hospital (2289 observations).

depending on the time/shift. Although we do not have a direct measure of hospital service quality, we have some information about the doctors attending the birth for a subsample of births. In table A5 we show that the number of doctors and the proportion of male doctors is balanced across different times of day. Additionally, we provide more systematic evidence in favor of our exclusion restriction by performing the analysis using variation in the probability of having a c-section only during the night, thus holding the quality of medical care constant (see section 5.1).

4. Results

Tables 2 and 3 present the results for the OLS estimation of equation (1) for the different measures of neonatal health. In table 2, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. All specifications include hospital and weekday fixed effects, the sample is restricted to single births, unplanned c-sections and vaginal deliveries, and we cluster standard errors at the hospital-shift level[13]. The results show that delivering via a c-section is associated with a significant decline of Apgar scores 1 and 5. Table 3 presents the results for other outcomes of neonatal health. As it can be seen, babies born by cesarean section are more likely to need reanimation and to go to the intensive care unit, but they are no more likely to die.

Table 2: OLS Results – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.528***	-0.524***	-0.419***	-0.219***	-0.219***	-0.142***
	(0.057)	(0.057)	(0.061)	(0.038)	(0.037)	(0.043)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the results of OLS regressions of Apgar scores 1 and 5, respectively, on an indicator for an unplanned cesarean birth. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: OLS Results – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Unplanned CS	0.137***	0.102***	0.081***	0.062***	-0.001	-0.005
	(0.016)	(0.014)	(0.014)	(0.014)	(0.002)	(0.003)
Mean of Y	0.060		0.082		0.004	
Observations	5783		5782		5783	
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the results of OLS regressions of different indicators of neonatal health on an indicator for an unplanned cesarean birth. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

As explained above, these estimates are likely to be biased because mothers giving birth by c-section and vaginally are not comparable, and because we cannot identify which kind of c-section is driving the results. The results for the IV estimation of the effects of non-medically indicated c-sections on Apgar scores 1 and 5 are shown in Table 4[14]. The first stage F-statistics are larger than 34 for the different specifications, so following Stock and Yogo (2005) critical values with one endogenous variable and one IV (16.38), we can reject the null hypothesis that our instrument is weak. In line with our descriptive analysis, Panel B shows that births that take place between 11 pm and 4 am are around 6 percentage points more likely to be by cesarean[15].

In the first row of the table below (Panel A), we observe that a c-section has a negative impact on both Apgar score 1 and

[13] All estimations hereafter use clustered standard errors at the hospital-shift level. We show in Table A6 in the appendix that our IV results are robust to alternative standard error estimations.

[14] The full regression output for both the first and second stage can be found in tables B1 and B2 in the appendix.

[15] We have also considered alternative specifications of the IV, using dummies for single hours in the window from 11 pm to 4 am. Our second stage results are similar but the first stage is weaker, thus harming precision and raising concerns about bias of the 2SLS. Results are available upon request.

Table 4: IV Estimation – Apgar Scores

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.122**	-1.147**	-0.992*	-0.956**	-0.987**	-0.936**
	(0.497)	(0.501)	(0.572)	(0.404)	(0.408)	(0.464)
Mean of Y		8.895			9.798	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: IV Estimation – Other Outcomes

	Intensive Care Unit		Reanimation		Neonatal death	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. 2SLS						
Unplanned CS	0.154	0.092	0.101	0.057	0.030	0.026
	(0.103)	(0.114)	(0.114)	(0.133)	(0.031)	(0.035)
Mean of Y	0.060		0.082		0.004	
Panel B. First stage						
Early night	0.073***	0.063***	0.073***	0.063***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5782	5782	5783	5783
First-stage F	41.591	34.234	41.576	34.149	41.591	34.234
Maternal controls	✓	✓	✓	✓	✓	✓
Pregnancy controls		✓		✓		✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on different indicators of neonatal health. The outcome variable in columns (1)-(2) is a dummy variable equal to one if the newborn was admitted to the intensive care unit; in columns (3)-(4), an indicator for whether the newborn needed reanimation (assisted ventilation), and in columns (5)-(6) an indicator of neonatal death. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling for maternal characteristics, weekday and hospital fixed effects; in the second column pregnancy controls are also added. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Apgar score 5. The estimated effects are large and significant. In the specification with the full set of controls (column 3), an unscheduled c-section reduces Apgar score 1 by 0.992 points. This effect is around 0.9 standard deviations (1.117) and is significant at the 10% significance level. A c-section also has a negative impact on Apgar score 5. In this case the coefficient is -0.936, larger than one standard deviation (0.818) and significant at the 5% significance level.

Most of the newborns in our sample have an Apgar score 1 equal to 9 and an Apgar score 5 equal to 10 (see figure A5). We thus perform a similar analysis but using as dependent variable an indicator for having Apgar scores 1 and 5, respectively, lower than 10 (table A7), and both scores lower than 9 (table A8). Our qualitative conclusions hold, as we find that a non-medically justified c-section, as compared to a vaginal delivery, increases the probability of having Apgar scores 1 and 5, respectively, below 10 by around 25 and 40 percentage points, and the probability of having Apgar scores 1 and 5 below 9 by 36 and 19 percentage points. Finally, Figure A6 in the appendix provides an overview of the size of the coefficients for different thresholds of Apgar 1 and 5, respectively, as dependent variables. This is relevant, since decreases in Apgar scores are non-linearly related to the health of the newborn. We see a clearer pattern for Apgar scores 5: there seems to be an effect of these non-medically justified interventions on the probability of having Apgar scores lower than 10, 9 and 8, but not lower than 7 or inferior levels. Therefore, these marginal c-sections increase the probability of deviating from the perfect scores, which are the mode in our sample, but we do not see significant effects in the left tail of the distribution.

We also perform the same analysis for other infant health outcomes. Results can be found in Table 5. Although we might expect an effect on needing intensive care, reanimation, or neonatal mortality, we do not observe any significant impact.

Our IV identifies the local average treatment effect for the "marginal" women, that is, for the deliveries that are sensitive to the subjective assessment of the doctor. More specifically, we capture cases in which the time of birth affects the decision of the doctor to perform a cesarean section. We therefore focus on c-sections that are not strictly necessary in the medical sense and that are potentially avoidable surgeries. These are, in fact, arguably the most relevant from a policy point of view. We are not able to estimate the effect for women who have a clear indication for a vaginal delivery or for women who receive c-sections that are medically indicated.

If we compare the results from the IV and OLS estimations, the IV coefficients are larger in absolute terms for Apgar scores. This can be explained by the fact that with the OLS estimation we include medically indicated c-sections, which reduce fetal distress and this partially offsets the negative effects of the non-medically indicated c-sections that we find when using our instrument.

However, if we compare the results for the other outcomes (see tables 3 and 5), we observe that in this case OLS coefficients are larger and significant: c-sections are associated with an increased probability of needing intensive care and reanimation. This suggests that these medically-indicated c-sections are performed in order to assist infants in distress who need immediate support. On the other hand, the IV estimates are not significant, arguably because the effects of non-medically indicated c-sections are short-lived: in spite of the worsening in Apgar scores, we do not find substantial evidence that these negative effects translate into needing intensive care, reanimation, or increased mortality risk.

To support the interpretation that our IV identifies the effect of non-medically indicated c-sections, we provide evidence that the c-sections captured by our instrument are not correlated with indications that should predict a medically necessary cesarean. In particular, we show that, while unplanned c-sections are in general strongly correlated with fetal distress, as measured by the level of intrapartum pH, we do not see any relationship when we focus on the predicted c-sections from our first stage. This comparison can be found in table A9 in the appendix.

So far, our analysis has compared c-sections with all vaginal births. The latter comprise two main categories: eutocic births – without any instrumentation – and operative (or instrumented) vaginal deliveries, which involve the use of forceps, vacuum or spatula. Medical studies have documented a negative association between operative vaginal deliveries and infant health (American College of Obstetricians and Gynecologists, 2015). Moreover, the decision to perform these procedures is also subject to variation at the provider level (Webb, 2002). For a cleaner comparison without the potential manipulation of the control group, we perform the same analysis comparing c-sections with eutocic deliveries. We would expect the effects of non-medically indicated c-sections to be stronger if compared with this group. The results in table A10 seem to confirm this hypothesis, and we also observe a slightly stronger first stage, suggesting that physician impatience might also lead to an increased use of instrumentation in the early hours of the night.

5. Robustness checks and extensions

5.1. Exclusion restriction: variation within the night

One potential concern of our identification strategy is that the quality of medical care could differ during the day compared to the night. Hence, it may be that the negative effects that we find on infant health are not due to the increased probability of having a c-section, but rather to a reduction in the quality of care during this time.

To further investigate this issue, we perform the same IV estimation but restricting the sample to mothers who gave birth during the night. We thus use variation in the probability of having a c-section during the night, holding the quality of care constant. As before, our instrument is an indicator variable equal to 1 if the woman gives birth during the early hours of the night (from 11 pm to 4 am). The sample is restricted to deliveries taking place from 8 pm to 8 am; i.e., during the last half of physicians' shifts, when healthcare professionals in the labor room – both obstetricians and midwives – do not change.

Results for the IV estimation using variation during the night can be found in Table 6. Despite the smaller sample size, we again find that a c-section reduces both Apgar scores 1 and 5. The coefficients remain large and significant, in particular so for Apgar 5. We interpret these results as evidence in favor of our exclusion restriction.

5.2. Excluding inductions

The comparison of maternal characteristics in Table 1 showed that mothers giving birth in the early hours of the night are more likely to have had their labor induced. Inductions can be scheduled, typically because the pregnancy has gone beyond full term and labor has not spontaneously started, or can be unscheduled if the mother's waters break but labor does not begin (NICE, 2008). If an induction is to be scheduled, the hospitals in our sample usually plan the latter for the morning, such that after progression of labor at average pace these women are expected to give birth in the evening or during the early hours of the night.

Table 6: IV Estimation – Apgar Scores during the Night

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.530*	-1.524*	-1.413	-1.511**	-1.512**	-1.535**
	(0.814)	(0.830)	(0.964)	(0.653)	(0.663)	(0.766)
Mean of Y		8.879			9.790	
Panel B. First stage						
Early Night	0.054***	0.053***	0.044***	0.053***	0.053***	0.044***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	3023	3023	3023	3022	3022	3022
First-stage F	17.217	16.619	12.812	17.144	16.537	12.760
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for births that took place between 8 pm and 8 am. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that took place during the night. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The relation between inductions and c-sections is a question where the medical literature and medical practice seem to differ. We observe in our sample that mothers with induced labor are more likely to have a c-section (see table A4). However, the recent medical literature finds that, while c-sections are conventionally regarded as the main potential complication of inductions, inductions at full term do not increase the risk of cesarean delivery (Saccone and Berghella, 2015) or even lower it (Mishanina et al., 2014), with no increased risks for the mother and some benefits for the fetus. All in all, it seems that whether or not a c-section is needed in cases of induced labor is likely to be dependent on the assessment of the obstetrician, such that mothers having had inductions probably fall into a "gray area" where we expect doctors' decisions to be more sensitive to external factors and incentives.

In any case, even if the decision to perform a c-section on mothers with induced labor was more dependent on doctors' routines or incentives than on the health conditions of the mother and the baby, if our analysis was driven by this type of mother alone, we would not be able to disentangle the effect of c-sections from the effect of medical inductions. In our main specifications we directly control for whether labor was induced, but in Table 7 we also repeat our analysis excluding inductions from our sample[16]. Here we see that, despite the reduction in the number of observations, our qualitative conclusions hold: births in the early night are still more likely to end

[16]The results for both the specification without inductions and the specification with only births during the night for reanimation, ICU admission, and neonatal death are consistent with those of table 5. Results are available upon request.

up as cesarean sections, and these have a negative and significant impact on Apgar scores. We thus conclude that, although inductions seem to make our first stage stronger as they might offer room for discretionary behavior, our findings do not depend on including them.

Table 7: Robustness Check – Excluding Inductions

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.747	-1.769	-1.804	-1.804*	-1.847*	-1.921*
	(1.086)	(1.104)	(1.171)	(0.931)	(0.952)	(1.011)
Mean of Y		8.952			9.828	
Panel B. First stage						
Early Night	0.037***	0.037***	0.035***	0.037***	0.037***	0.035***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	4369	4369	4369	4367	4367	4367
First-stage F	10.720	10.663	10.179	10.677	10.614	10.319
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, for non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, and an indicator for preterm birth. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) that were not induced. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.3. Falsification test

In order to lend support to the credibility of our identification strategy, we run additional "placebo" regressions using an outcome variable that is predetermined when the mother goes into labor, and thus should not be affected by a c-section. In particular, we analyze birth weight and weeks of gestation. The results of this analysis are reported in Table 8. As in previous tables, the first column for each outcome presents the results without controls, the second column incorporates controls for maternal characteristics, and finally the third column adds information about the pregnancy. The results of this exercise suggest that there is no effect of c-sections on birth weight or gestational weeks. This provides further evidence in favor of our specification.

5.4. Time of admission and time of birth

One potential concern with using time of birth as an instrument for the mode of delivery is that, given that cesarean sections by definition shorten labor, the exact time of birth will be influenced by the type of birth itself. In other words, one might be worried about reverse causality in the first stage. We argue that any potential bias should be alleviated by the specification of the instrument not as the time of birth itself, but as a relatively wide time interval (in particular, as a dummy equal to one for births between 11 pm and 4 am). Because the instrument is defined in this way, we do not need to assume that the exact time of birth is not influenced by the mode of delivery; it suffices that any impact of the decision about the type of birth on the time interval in which the delivery takes place is negligible.

Table 8: Placebo Regressions: Birth Weight and Gestational Weeks

	Birth Weight (in logs)			Gestational weeks		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-0.023	-0.027	0.042	0.250	0.203	0.081
	(0.077)	(0.076)	(0.077)	(0.774)	(0.772)	(0.866)
Mean of Y		8.080			39.266	
Observations	5782	5782	5782	5783	5783	5783
First-stage F	41.627	41.559	34.222	41.661	41.591	35.154
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on birth weight (in natural logs) and gestational weeks, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth (except in the regression of gestational weeks), and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In our context, if doctors' incentive is to perform a cesarean section to ongoing deliveries early at night that they expect to end up during their shift, it will likely be to mothers that are advanced in labor. Therefore, the counterfactual to the cesarean is expected to be a vaginal birth two or three hours later[17]; that is, for most c-sections in the early night, the counterfactual vaginal birth would have probably taken place in the early hours of the night as well. As a result, the change in the probability of giving birth between 11 pm and 4 am caused by having a c-section is expected to be small.

In order to assess empirically the magnitude of the potential bias, we use information about the time of admission of mothers to the hospital, which is only available for one of the hospitals in our sample. In particular, we want to see if our results are robust to substituting our instrument with one based on the time of admission. This alternative instrument should remove concerns about reverse causality since, for unscheduled deliveries, time of admission should not be affected by mode of delivery.

First, we explore the distribution of the c-section rate as a function of time of admission (see figure A7) and find that there is a similar peak to that in figure 1, in this case for mothers admitted between 2 pm and 8 pm. Therefore, we define our new instrument to be equal to one for mothers admitted during this time interval[18]. Results using this new instrument can be found in table 9, which follows the usual table structure. Panel B displays the coefficients of the first-stage regressions: in the third column for each outcome, which shows the results of the specification with the full set of controls, we can see that mothers that arrived at the hospital between 2 pm and 8 pm were around 6.3 percentage points more likely to have a c-section. This is the same result we found for mothers giving birth between 11 pm and 4 am: they are also 6.3 percentage points more likely

[17]See an explanation of the average time of each stage of labor in section 2.1.

[18]Following the same logic as in our main analysis, we select the interval in which the c-section rate is above 15%.

Table 9: Robustness check – IV Estimation with Admission Time Instrument

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.554**	-1.568*	-1.601*	-0.802	-0.791	-0.793
	(0.787)	(0.815)	(0.960)	(0.578)	(0.601)	(0.712)
Mean of Y		8.861			9.869	
Panel B. First stage						
Admission time 2pm-8pm	0.077***	0.074***	0.064***	0.077***	0.074***	0.063***
	(0.022)	(0.022)	(0.021)	(0.022)	(0.022)	(0.021)
Observations	2289	2289	2289	2287	2287	2287
First-stage F	12.079	11.601	9.465	12.029	11.550	9.423
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned c-section on Apgar scores 1 and 5, respectively. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for mothers admitted to the hospital between 2 pm and 8 pm. Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to have a cesarean birth. Panel A shows the 2SLS coefficients: despite the reduced sample size, we find very similar point estimates to those in table 4. The resemblance of these results to those in our main analysis suggests that reverse causality, in practice, does not have a large influence in our setting, and supports the validity of our instrument.

5.5. Another measure of neonatal health: umbilical cord pH

In addition to Apgar scores, reanimation, ICU admission and neonatal death, we also study the impact of cesarean sections on the pH of the umbilical cord. Although it has not been used in the economics literature, this measure of neonatal health has been widely analyzed in medical studies, and it is considered to add objective information to the Apgar score regarding the status of the newborn. Due to its objective nature, it is used to support medico-legal claims (Skiold et al., 2017). As explained in Section 3.1, the examination of the umbilical artery provides a measure of fetal distress. Although the relationship between pH levels and Apgar scores is not one-to-one, they are positively correlated[19]. The medical literature recommendation is to consider pH levels together with Apgar scores in order to assess the well-being of the newborn (Hannah, 1989; Malin et al., 2010).

Table 10 shows the results from the estimation of the impact of a c-section on the probability of the pH level being below different thresholds (7.20, 7.15 and 7.10) for the different samples: the full specification (columns 1–3), during the night (columns 4–6) and excluding inductions (7–9). This outcome was only recorded in 3 out of the 4 hospitals in our sample, and thus the number of observations is lower. All our estimates go in the same direction: c-sections increase the probability of pH levels being below the different thresholds, suggesting the presence of a negative health effect as measured by this outcome. The most consistent results are found for the pH threshold of 7.15. Our first stage F-statistic is strong for the full specification (25.58) but becomes weaker as the sample drops. Overall, these findings go in line with the previous results of a negative effect of c-sections on neonatal health.

Table 10: IV estimation — Umbilical cord pH level

	Full Specification			During the Night			Excluding Inductions		
pH threshold	7.20	7.15	7.10	7.20	7.15	7.10	7.20	7.15	7.10
Panel A. 2SLS									
Unplanned CS	0.303	0.341*	0.184	1.074*	0.857**	0.307	1.004	0.947*	0.573*
	(0.250)	(0.192)	(0.122)	(0.562)	(0.415)	(0.220)	(0.671)	(0.538)	(0.333)
Mean of Y	0.221	0.102	0.042	0.212	0.100	0.044	0.216	0.096	0.039
Panel B. First stage									
Early Night		0.063***			0.042***			0.033***	
		(0.012)			(0.014)			(0.012)	
Observations		4444			2316			3403	
First-stage F		25.589			8.567			6.992	

Notes: The table shows the instrumental variable estimates of the effect of an unplanned cesarean birth on the probability of the umbilical cord pH being below different thresholds (7.20, 7.15, and 7.10), for different samples. Columns (1)-(3) use the usual full sample, columns (4)-(6) use only births during the night, and columns (7)-(9) include only non-induced births. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor (except in the last three columns). Mean of Y refers to the average of the outcome variable in the sample. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6. Conclusions

This paper provides new credible evidence of the adverse effects of avoidable cesarean sections on newborn health. In order to overcome potential omitted variable bias and abstract from those cases in which c-sections respond to a clear clinical indication, we make use of a novel instrument that exploits variation in the probability of receiving a c-section that is unrelated to maternal and fetal health: variation in time of birth. Specifically, we document an increase in unplanned c-sections during the early hours of the night (from 11 pm to 4 am) that is not driven by different characteristics of mothers who give birth during this time, providing us with exogenous variation in the probability of the delivery ending up in a cesarean section.

Our findings suggest that these non-medically indicated c-sections lead to a significant worsening of newborn health, as measured by Apgar scores. According to the medical literature, deterioration in these outcomes might be capturing increased respiratory problems and reduced excitability and muscle tone (Hyde et al., 2012). However, the magnitude of our estimates suggests that these c-sections lead to a decrease of just around one point in Apgar scores 1 and 5 in otherwise healthy babies – the mean Apgar scores 1 and 5 are 8.9 and 9.8, respectively. Our analysis by thresholds of Apgar scores confirms that the effects of these c-sections are limited to the higher levels of these scales; in particular, we see an increased probability of having Apgar score 5 below 10, 9 and 8. It is worth noting that previous studies find worse long-run outcomes for newborns with these levels of Apgar, compared to their siblings with perfect scores, even if these levels are not generally considered to be concerning: Oreopoulos et al. (2008) find that individuals with

[19]Figure A8 in the appendix shows the distributions of umbilical cord pH for infants with Apgar scores 1 above and below 9 (first panel), and for infants with Apgar scores 5 above and below 9 (second panel). We observe that the distribution of pH levels for infants with Apgar scores below 9 is shifted to the left compared to that for babies with higher scores, with this being more salient for Apgar score 5.

Apgar scores of 7 or 8 are more likely to drop out or repeat a grade, and that those with Apgar scores between 7 and 9 are also more likely to receive social assistance after age 18.

In any case, we do not find evidence that these effects translate into a significant increase in the need for reanimation or intensive care, or into increased risk of neonatal death, which is consistent with the absence of significant impacts on lower levels of Apgar scores and on low thresholds of the pH of the umbilical cord. We can thus rule out very severe impacts at birth, as well as any short-run health benefit of these avoidable interventions. This is an important contribution, given that previous studies in the medical literature documented an association between c-sections and an increased risk of serious respiratory morbidity and subsequent admission to neonatal ICU (Grivell and Dodd, 2011). Their findings are consistent with the results of our OLS estimation, suggesting that former analysis might have been capturing the underlying health status of newborns who need a medically necessary cesarean.

However, it should also be pointed out that some effects of c-sections may not be visible at birth. In particular, medical studies suggest that the exposure of newborns to the maternal vaginal microbiota is interrupted with cesarean birthing, and that this could translate into increased risk for immune and metabolic disorders in the long run (Hyde et al., 2012; Dominguez-Bello et al., 2016). Any such effect need not be reflected in any of the short-run outcomes we are able to explore in this study, which limits the conclusions we can derive from our analysis. In this paper, however, we propose a new instrument that will make possible to examine this and other channels and gather evidence to obtain a more complete understanding of the causal effect of non-medically indicated c-sections on the health of the infant and the mother in the longer run.

Our results also highlight non-financial incentives as an important factor influencing the decision-making of health care providers. Although more work is needed to clearly understand the decisions of doctors driving the observed time variation in c-section rates, we have provided some suggestive evidence that stresses the potential role of leisure incentives in the context of public hospitals, and which is consistent with the findings of previous studies. In particular, our findings suggest that doctors may be less tolerant to the time-consuming natural progression of labor during times of day when leisure incentives are more salient, and thus are more willing to perform procedures that accelerate the delivery. Along this line, our results point to the need to revise the incentives created by the shift structure and long working hours of physicians, so as to reduce avoidable interventions.

A simple back-of-the-envelope calculation can shed some light on the potential gains that could result from such reduction. The first-stage coefficient from our main specification with all controls (column 3 in table 4) implies that, holding all other characteristics constant, during the early hours of the night the c-section rate increases by 6.3 percentage points compared to the rest of the day. Given that the c-section rate in our sample of hospitals is 16.5%, removing these excess c-sections would lower the c-section rate by 38.1% – or equivalently, a decrease of 245 c-sections per year[20]. Taking into account that the average cost of a c-section for the Spanish public health system is 1692.97 Euros higher than that of a vaginal delivery[21], by cutting these excessive c-sections, hospitals in our sample could achieve a cost reduction of around 675,500 Euros. Applying the same logic for all births that took place in Spanish public hospitals in 2014, this would result in savings of more than 47 million Euros for the Spanish health system[22]. To give some meaning to these numbers, given that the average annual salary for a speciality doctor is 45,970 Euros[23] and there are 453 public hospitals in Spain, these savings would enable each hospital to hire more than 2 additional doctors. An increase in the number of obstetricians could help, in turn, to alleviate the need for such long working hours. Importantly, these savings could be materialized without harming neonatal health, given the absence of benefits of these avoidable c-sections.

Adecco Healthcare, 2017. Empleo en el sector sanitario. Adecco Group, https://adecco.es/wp-content/uploads/2017/07/NdP-Las-ofertas-de-empleo-en-el-sector-sanitario-se-incrementarán-un-40-según-Adecco-Healthcare.pdf. Accessed: 26-12-2017.

AEPED, 2014. Test de apgar. Asociación Española de Pediatría (Spanish Pediatrics Association), http://www.aeped.es/rss/en-familia/test-apgar. Accessed: 20-12-2017.

Almond, D., Chay, K. Y., Lee, D. S., 2005. The costs of low birth weight. The Quarterly Journal of Economics 120 (3), 1031–1083.

Almond, D., Currie, J., 2011. Killing me softly: The fetal origins hypothesis. Journal of Economic Perspectives 25 (3), 153–72.

American College of Obstetricians and Gynecologists, 2006. Mode of Term Singleton Breech Delivery. ACOG Commitee Opinion No. 340. Obstetrics and Gynecology 108, 235–237.

American College of Obstetricians and Gynecologists, 2015. Practice Bulletin No. 154: Operative Vaginal Delivery. Obstetrics and Gynecology 126 (5), e56–e65.

APA, 2017. Cesarean procedure. American Pregnancy Association, http://americanpregnancy.org/labor-and-birth/cesarean-procedure/. Accessed: 11-12-2017.

Baicker, K., Buckles, K. S., Chandra, A., 2006. Geographic variation in the appropriate use of cesarean delivery. Health Affairs 25 (5), w355–w367.

Barber, E. L., Lundsberg, L. S., Belanger, K., Pettker, C. M., Funai, E. F., Illuzzi, J. L., Jul 2011. Indications Contributing to the Increasing Cesarean Delivery Rate. Obstetrics and Gynecology 118 (1), 29–38.

Boksa, P., Zhang, Y., 2008. Epinephrine administration at birth prevents long-term changes in dopaminergic parameters caused by cesarean section birth in the rat. Psychopharmacology 200 (3), 381–391.

Brown, H., 1996. Physician demand for leisure: implications for cesarean section rates. Journal of Health Economics 15 (2), 233 – 242.

Currie, J., MacLeod, W. B., 2008. First Do No Harm? Tort Reform and Birth Outcomes. Quarterly Journal of Economics 123 (2), 795–830.

Davidson, R., Roberts, S. E., Wotton, C. J., Goldacre, M. J., 2010. Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. BMC Pulmonary Medicine 10 (1), 14.

[20] This figure is calculated with data from 2015, when there were 644 cesareans out of 4027 births in the four hospitals of our sample.

[21] The Spanish National Health System estimated that, for the year 2014, the average cost of a cesarean section without complications was 3,739.06 Euros, while that of a vaginal birth without complications was 2,046.09 Euros. See Ministerio de Sanidad, Servicios Sociales e Igualdad (2014).

[22] The c-section rate for all public hospitals in Spain in 2014 was 22.1%. Assuming that these hospitals have a similar time variation in the c-section rate, removing the excessive c-sections of the early hours of the night would result in a c-section rate of 13.68%. Given that there were 332,252 births, the number of c-sections would decrease from 73,411 to 45,452; that is, a reduction of 27,959 c-sections per year.

[23] Adecco Healthcare (2017)

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Hoashi, M., Rivera-Vinas, J. I., Mendez, K., Knight, R., Clemente, J. C., 2016. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nature Medicine 22, 250–253.

Figlio, D., Guryan, J., Karbownik, K., Roth, J., 2014. The effects of poor neonatal health on children's cognitive development. American Economic Review 104 (12), 3921–55.

Fraser, W., Usher, R. H., McLean, F. H., Bossenberry, C., Thomson, M. E., Kramer, M. S., Smith, L., Power, H., 1987. Temporal variation in rates of cesarean section for dystocia: Does "convenience" play a role? American Journal of Obstetrics and Gynecology 156 (2), 300 – 304.

Grant, D., 2009. Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. Journal of Health Economics 28 (1), 244 – 250.

Grivell, R. M., Dodd, J. M., 2011. Short- and long-term outcomes after cesarean section. Expert Review of Obstetrics and Gynecology 6 (2), 205–215.

Halla, M., Mayr, H., Pruckner, G. J., Garcia-Gomez, P., 2016. Cutting fertility? the effect of cesarean deliveries on subsequent fertility and maternal labor supply. IZA Discussion Paper.

Hannah, M., February 1989. Birth asphyxia: does the Apgar score have diagnostic value? Obstetrics and Gynecology 73 (2), 299—300.

Hansen, A. K., Wisborg, K., Uldbjerg, N., Henriksen, T. B., 2008. Risk of respiratory morbidity in term infants delivered by elective caesarean section: cohort study. BMJ : British Medical Journal 336, 85–87.

Hueston, W. J., McClaflin, R. R., Claire, E., 1996. Variations in cesarean delivery for fetal distress. The Journal of Family Practice 43 (5), 461–467.

Hyde, M. J., Mostyn, A., Modi, N., Kemp, P. R., 2012. The health implications of birth by caesarean section. Biological Reviews 87 (1), 229–243.

Jachetta, C., 2015. Cesarean sections and later health outcomes. http://www.christinejachetta.com/upload/jachetta_job_market_paper_2015.pdf. Accessed: 04-10-2016.

Jensen, V. M., Wüst, M., 2015. Can caesarean section improve child and maternal health? the case of breech babies. Journal of Health Economics 39, 289 – 302.

Klasko, S. K., Cummings, R. V., Balducci, J., DeFulvio, J. D., Reed, J. F., 1995. The impact of mandated in-hospital coverage on primary cesarean delivery rates in a large nonuniversity teaching hospital. American Journal of Obstetrics and Gynecology 172 (2), 637 – 642.

Lefèvre, M., 2014. Physician induced demand for c-sections: does the convenience incentive matter? Health, Econometrics and Data Group (HEDG) Working papers, HEDG, c/o Department of Economics, University of York.

Malin, G. L., Morris, R. K., Khan, K. S., 2010. Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis. BMJ: British Medical Journal 340, c1471.

Marcos, J. C. M., 2008. Cesarea a Demanda. Sociedad Española de Ginecología y Obstetricia.

Mühlrad, H., 2017. Cesarean section for high-risk births: short and long term consequences for breech births. https://www.dropbox.com/s/xto1qlsf1qtxf7d/C-section_Muhlrad.pdf. Accessed: 26-12-2017.

Ministerio de Sanidad, Servicios Sociales e Igualdad, 2014. Costes hospitalarios - contabilidad analítica. Spanish Ministry of Health, Social Services and Equality, http://www.msssi.gob.es/estadEstudios/estadisticas/inforRecopilaciones/anaDesarrolloGDR.htm. Accessed: 13-01-2017.

Ministerio de Sanidad, Servicios Sociales e Igualdad, 2015. Informe anual del sistema nacional de salud. Spanish Ministry of Health, Social Services and Equality, http://www.msssi.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnSNS.htm. Accessed: 13-01-2017.

Ministerio de Sanidad y Política Social, 2009. Informe maternidad hospitalaria: Estándares y recomendaciones. Spanish Ministry of Health and Social Policy, http://www.msc.es/organizacion/sns/planCalidadSNS/docs/AHP.pdf. Accessed: 14-12-2017.

Ministerio de Trabajo y Asuntos Sociales, 1997. BOE-A-1997-26857. Https://www.boe.es/diario_boe/txt.php?id=BOE-A-1997-26857.

Mishanina, E., Rogozinska, E., Thatthi, T., Uddin-Khan, R., Khan, K. S., Meads, C., 2014. Use of labour induction and risk of cesarean delivery: a systematic review and meta-analysis. Canadian Medical Association Journal 186 (9), 665–673.

NICE, 2008. Inducing labour. NICE guideline CG70, National Institute for Health and Care Excellence, https://www.nice.org.uk/guidance/cg70. Accessed: 08-06-2017.

NICE, 2011. Cesarean section. NICE clinical guideline CG190, National Institute for Health and Care Excellence, https://www.nice.org.uk/guidance/cg132/chapter/1-guidance. Accessed: 06-12-2017.

NICE, 2014. Intrapartum care for healthy women and babies. NICE guideline CG190, National Institute for Health and Care Excellence, https://www.nice.org.uk/guidance/cg190. Accessed: 06-02-2017.

NICE, 2016. Deciding whether to offer caesarean section. NICE pathways, National Institute for Health and Care Excellence, http://pathways.nice.org.uk/pathways/caesarean-section. Accessed: 04-10-2016.

OECD, 2013. Health at a Glance 2013: OECD Indicators. OECD Publishing, Paris.

Oreopoulos, P., Stabile, M., Walld, R., Roos, L. L., 2008. Short-, medium-, and long-term consequences of poor infant health: An analysis using siblings and twins. Journal of Human Resources 43, 88–138.

Otamiri, G., Berg, G., Ledin, T., Leijon, I., Lagercrantz, H., 1991. Delayed neurological adaptation in infants delivered by elective cesarean section and the relation to catecholamine levels. Early Human Development 26 (1), 51 – 60.

Saccone, G., Berghella, V., 2015. Induction of labor at full term in uncomplicated singleton gestations: a systematic review and metaanalysis of randomized controlled trials. American Journal of Obstetrics and Gynecology 213 (5), 629 – 636.

SEGO, 2005. Monitorización fetal intraparto. Progresos de Obstetricia y Ginecología 48 (4), 207–216.

Sevelsted, A., Stokholm, J., Bønnelykke, K., Bisgaard, H., 2015. Cesarean section and chronic immune disorders. Pediatrics 135 (1), e92–e98.

Shurtz, I., 2013. The impact of medical errors on physician behavior: Evidence from malpractice litigation. Journal of Health Economics 32 (2), 331 – 340.

Skiold, B., Petersson, G., Ahlberg, M., Stephansson, O., Johansson, S., 2017. Population-based reference curve for umbilical cord arterial pH in infants born at 28 to 42 weeks. Journal of Perinatology 37 (3), 254–259.

Spetz, J., Smith, M. W., Ennis, S. F., 2001. Physician incentives and the timing of cesarean sections: evidence from California. Medical care 39 (6), 536–550.

Spong, C. Y., Berghella, V., Wenstrom, K. D., Mercer, B. M., Saade, G. R., 2012. Preventing the first cesarean delivery: Summary of a joint Eunice Kennedy Shriver National Institute of Child Health and Human Development, Society for Maternal-fetal Medicine, and American College of Obstetricians and Gynecologists workshop. Obstetrics and Gynecology 120 (5).

Stock, J., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D. W. (Ed.), Identification and Inference for Econometric Models. Cambridge University Press, New York, pp. 80–108.

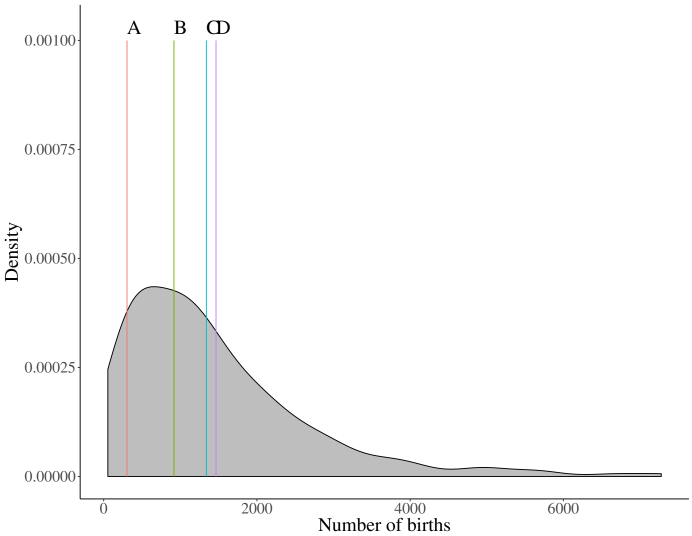
Webb, D. A., 2002. Time of day variation in rates of obstetric intervention to assist in vaginal delivery. Journal of Epidemiology & Community Health 56 (8), 577–578.

WHO, 2015. WHO Statement on Caesarean Section Rates. World Health Organization, Geneva.

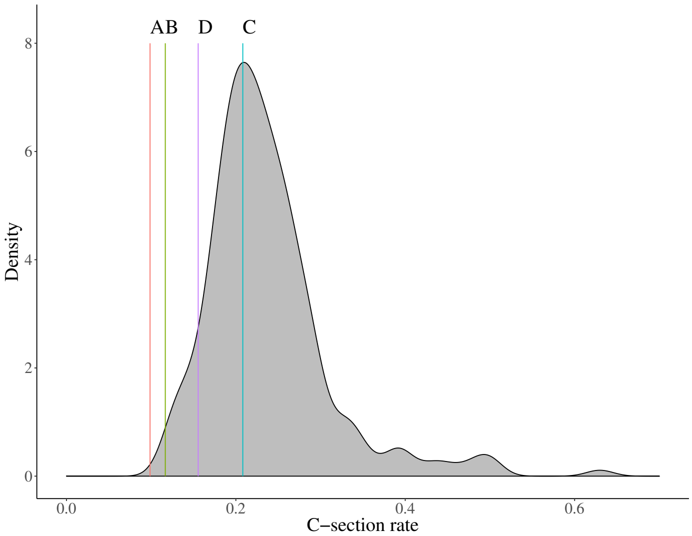
Zanardo, V., Simbi, A., Franzoi, M., Soldá, G., Salvadori, A., Trevisanuto, D., 2004. Neonatal respiratory morbidity risk and mode of delivery at term: influence of timing of elective caesarean delivery. Acta Paediatrica 93 (5), 643–647.

Appendix

Figure A1: Distribution of Number of Births and C-Section Rates in all Spanish Public Hospitals



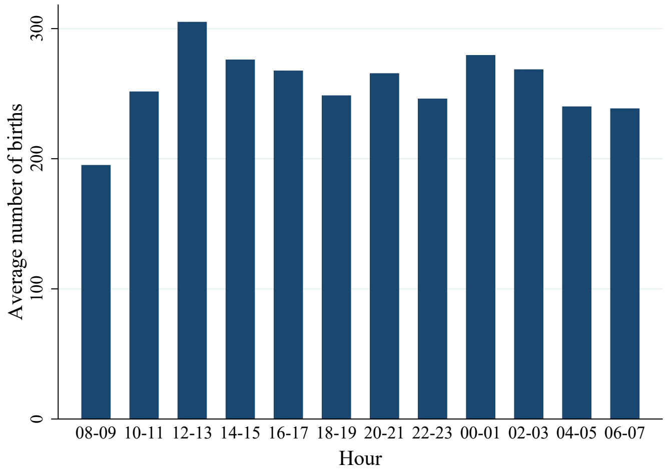
(a) Number of Births Attended in a Year



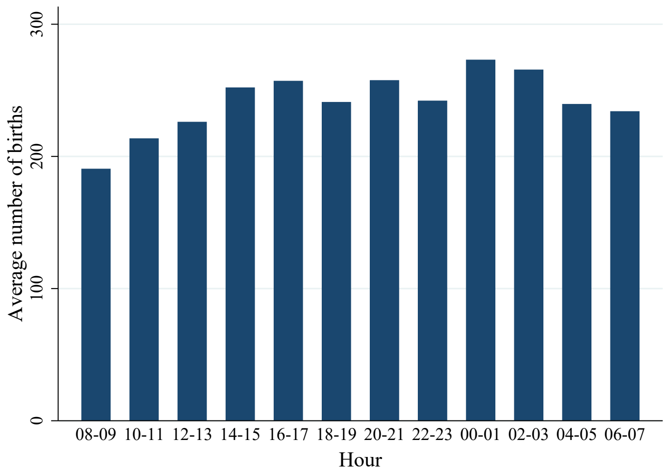
(b) C-Section Rate in a Year

Notes: Figure (a) shows the distribution of the number of births attended in one year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Figure (b) shows the distribution of c-section rates in a year for all Spanish Public Hospitals compared to hospitals in our sample (A, B, C and D). Source: our data (2015) and Estadística de Centros Sanitarios de Atención Especializada (2013).

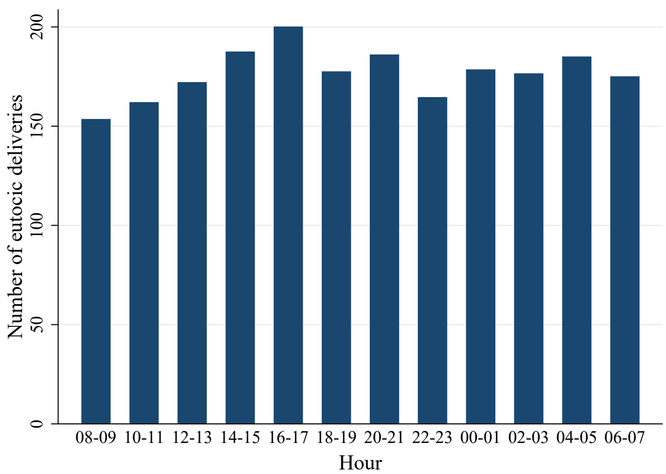
Figure A2: Distribution of Different Types of Births across Times of Day



(a) All Births



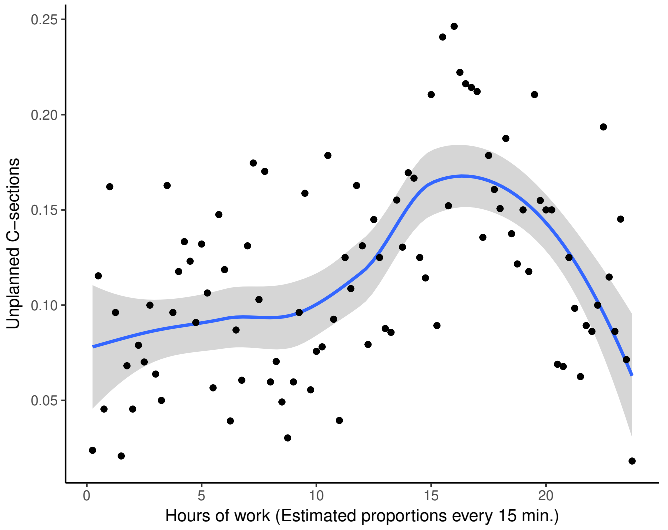
(b) Unplanned C-Sections and Vaginal Births



(c) Eutocic Deliveries

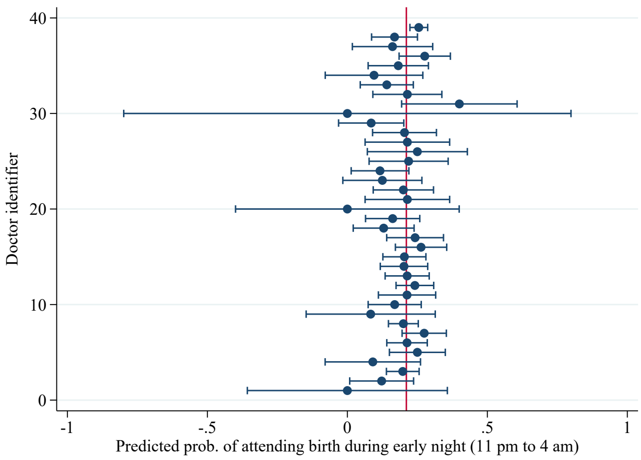
Notes: These figures represent the distribution of different types of births across times of day, grouped by intervals of two hours. Figure (a) represents the number of births per two hours using the full sample of 6,163 observations. Figures (b)-(c) use our usual sample of 5,783 observations. Figure (b) shows the number of births per two hours in this restricted sample, which includes only unplanned c-sections or vaginal births (excluding breech vaginal births), while figure (c) displays the number of eutocic deliveries.

Figure A3: Proportion of Unplanned C-Sections by Physicians' Hours Worked (Loess Estimate)



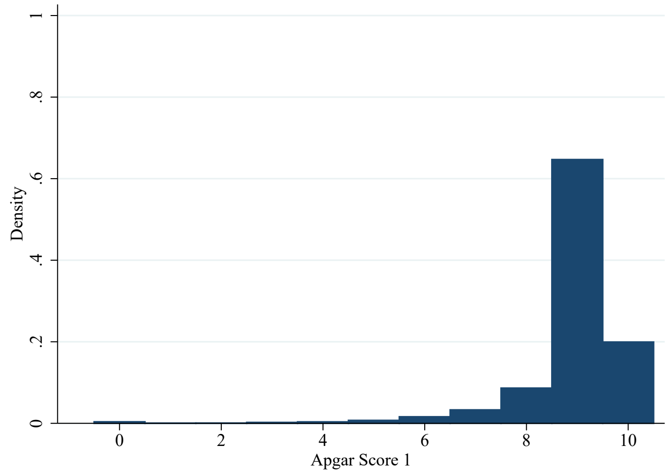
Notes: This figure shows the LOESS or local regression estimate of the proportion of observed unplanned c-sections as a function of a 24h shift, starting at 8 am and finishing at 8 am of the following day with a span of 15 minutes. The shaded area shows the 95% confidence interval.

Figure A4: Predicted Probability by Doctor of Attending Births during the early hours of the night

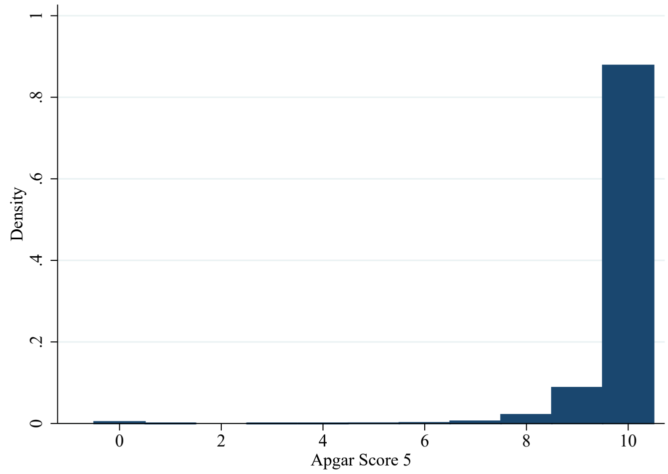


Notes: The figure shows the probability of attending births during the early hours of the night across different doctors, for a subsample of births for which the doctor identifier was registered (N=3,018). Sample is further restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Figure A5: Distribution of Apgar Scores



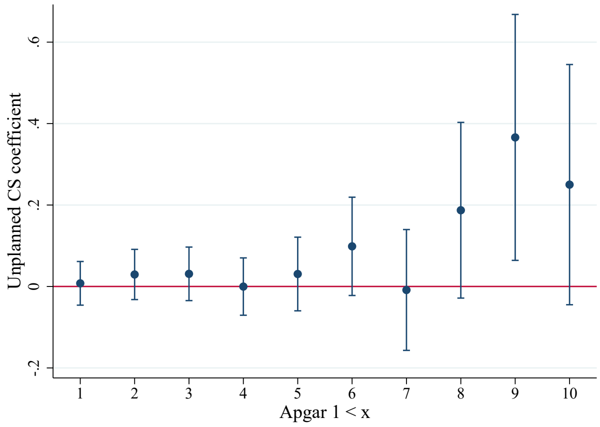
(a) Apgar Score 1



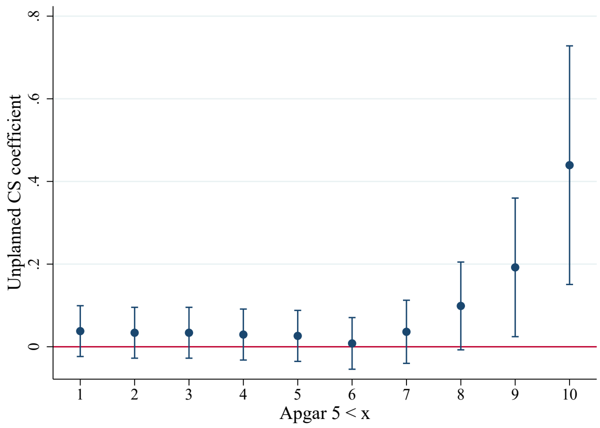
(b) Apgar Score 5

Notes: These figures show the distribution of Apgar scores for all births. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Figure A6: IV Coefficients by Apgar Threshold



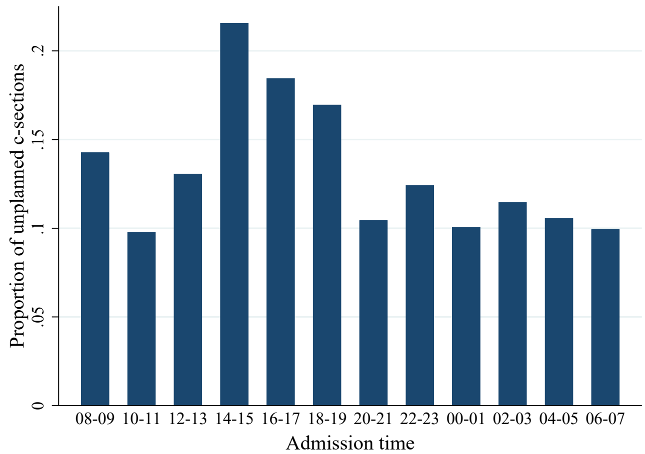
(a) Apgar Score 1



(b) Apgar Score 5

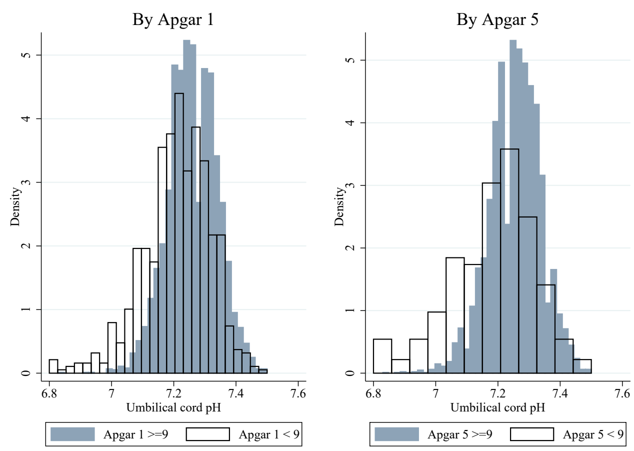
Notes: The figures show the second stage coefficients for the IV regressions of the effect of an unplanned c-section on the probability of Apgar scores being below different thresholds, in regressions with the full set of pregnancy and maternal controls. Figure (a) shows the coefficients for Apgar score at minute 1 after birth. Figure (b) shows the coefficients for Apgar score at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Figure A7: Proportion of Unplanned C-Sections by Time of Admission



Notes: The figure shows the proportion of unplanned c-sections over the sample of unplanned c-sections and vaginal births, by time of admission to the hospital. Sample is restricted to one hospital (C), single births, unscheduled c-sections and vaginal births (excluding breech babies).

Figure A8: Distribution of Umbilical Cord pH by Levels of Apgar 1 and 5



Notes: These figures show the distribution of values of umbilical cord pH by Apgar scores above or below 9. Figure (a) shows the distribution for Apgar scores at minute 1 after birth. Figure (b) shows the distribution for Apgar scores at minute 5 after birth. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies)

Table A1: Summary Statistics

	Mean	SD
A. Mother characteristics		
Mother's age	31.890	5.414
Level of education		
No school	0.032	0.175
Primary school	0.257	0.437
Secondary school	0.523	0.500
University education	0.188	0.391
Non-Spanish	0.250	0.433
Single	0.017	0.130
Mother's weight	65.715	14.536
Mother's height	1.638	2.087
B. Pregnancy characteristics		
Tobacco during pregnancy	0.122	0.327
Alcohol during pregnancy	0.004	0.062
Previous c-section	0.113	0.317
Gestation weeks	39.204	1.785
Multiple pregnancy	0.004	0.064
Obstetric Risk	0.406	0.491
Induction	0.227	0.419
C. Type of birth		
Planned c-section	0.053	0.224
Unplanned c-section	0.112	0.316
Spatula	0.007	0.084
Eutocic	0.687	0.464
Forceps	0.0141	0.118
Breech Vaginal	0.001	0.036
Vacuum	0.125	0.331
D. Newborn outcomes		
Apgar 1	8.884	1.117
Apgar 5	9.793	0.818
Birth weight (in gr.)	3267.970	519.988
Low birth weight (<2500 gr.)	0.068	0.252
Intensive care unit	0.064	0.244
Reanimation	0.084	0.277
Neonatal death	0.004	0.061
Umbilical cord pH	7.254	0.086
Intrapartum pH	7.273	0.073
Male	0.521	0.500
Observations	6163	

Notes: The table shows means and standard deviations for the outcome variables and a set of background variables for all births in our sample of public hospitals.

Table A2: Summary Statistics of All Births in Spanish Hospitals (2014-2015)

	Mean	SD
Mother's age	32.274	5.449
Non-Spanish	0.180	0.384
Gestation weeks	39.024	1.919
Multiple pregnancy	0.023	0.149
Birth weight (in gr.)	3227.344	531.320
Low birth weight (<2500 gr.)	0.069	0.253
Male	0.516	0.500
Observations	827,692	

Notes: This table shows descriptive statistics from all births in Spanish hospitals in 2014 and 2015. Source: Spanish National Statistics Institute, births microdata.

Table A3: First Stage: Busy vs. Non-Busy Nights

	(1) Single-birth nights	(2) Multiple-birth nights
Early Night	0.092***	0.054***
	(0.023)	(0.012)
Observations	1471	3733

Notes: The table shows the results of the first stage estimation on two different samples: single and multiple birth nights. The coefficients are OLS estimates of the regression of an indicator for an unplanned cesarean birth on an indicator for births during the early hours of the night (from 11 pm to 4 am). Single-birth nights are defined as days in which there is only one delivery from 8 pm to 8 am, whereas multiple-birth nights are those in which more than one delivery occurs during these times. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is in all cases restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Maternal Characteristics by Type of Birth

	Means		p-value
	Vaginal birth	Unplanned CS	for difference
A. Personal characteristics			
Mother's age	31.622	32.828	0.000
Level of education			
No school	0.033	0.022	0.126
Primary school	0.263	0.206	0.001
Secondary school	0.514	0.609	0.000
University education	0.191	0.164	0.083
Non-Spanish	0.255	0.199	0.001
Single	0.017	0.015	0.662
Mother's weight	65.312	67.830	0.000
Mother's height	1.646	1.595	0.559
B. Pregnancy characteristics			
Tobacco during pregnancy	0.120	0.134	0.277
Alcohol during pregnancy	0.003	0.007	0.089
Gestation weeks	39.320	38.863	0.000
Previous c-section	0.076	0.223	0.000
Obstetric risk	0.367	0.580	0.000
Intrapartum pH	7.288	7.245	0.000
Birth weight	3288.492	3181.038	0.000
Induction	0.214	0.431	0.000
Observations	5098	685	5783

Notes: The table shows means for a set of maternal and pregnancy characteristics by type of birth and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech babies).

Table A5: Doctor Characteristics by Time of Day

	Means		p-value
	Not early night	Early night	for difference
Male doctor	0.205	0.217	0.538
Number of doctors	1.568	1.603	0.286
Observations	1827	511	2338

Notes: The table shows the mean proportion of male doctors and number of doctors by time of day and the p-value for the difference between the means of the two groups. Sample is restricted to single births, unscheduled c-sections and vaginal births (excluding breech vaginal babies).

Table A6: IV Estimation – Apgar Scores: Standard Errors Robustness

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Unplanned CS	-0.992*	-0.992*	-0.992*	-0.936**	-0.936**	-0.936**
	(0.577)	(0.572)	(0.568)	(0.461)	(0.464)	(0.465)
Mean of Y		8.895			9.798	
Observations		5783			5781	
Cluster (shift)	✓			✓		
Cluster (hospital-shift)		✓			✓	
Robust			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, comparing alternative standard error estimations. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). The first column for each outcome has clustered standard errors at the shift level; in the second column standard errors are clustered at the hospital-shift level, as in our main specification, and in the third column we estimate heteroscedasticity-robust standard errors. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: IV Estimation – Apgar Score < 10

	Apgar Score 1 <10			Apgar Score 5 <10		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.283*	0.285*	0.250	0.433***	0.445***	0.439***
	(0.157)	(0.158)	(0.182)	(0.146)	(0.147)	(0.170)
Mean of Y		0.801			0.122	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 10. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8: IV Estimation – Apgar Score < 9

	Apgar Score 1 <9			Apgar Score 5 <9		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	0.380**	0.391**	0.366**	0.189**	0.192**	0.192*
	(0.158)	(0.159)	(0.183)	(0.088)	(0.089)	(0.103)
Mean of Y		0.154			0.034	
Panel B. First stage						
Early night	0.073***	0.073***	0.063***	0.073***	0.073***	0.063***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	5783	5783	5783	5781	5781	5781
First-stage F	41.661	41.591	34.234	41.570	41.487	34.159
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on the probability of Apgar scores 1 and 5, respectively, being lower than 9. The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies). Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Robustness Check: Fetal Distress and C-Sections

	Unplanned CS		Predicted CS	
	(1)	(2)	(1)	(2)
Intrapartum pH	-1.768***		0.018	
	(0.281)		(0.019)	
Intra. pH < 7.2		0.312***		-0.002
		(0.060)		(0.004)
Observations	425	425	425	425

Notes: The table shows the results of OLS regressions of all unplanned cesarean sections and the time-predicted c-sections on indicators of fetal distress. In the first two columns the dependent variable is an indicator equal to one for all unplanned c-sections, while in the last two columns the dependent variable takes the fitted values from the first-stage regression. In the first column for each outcome the explanatory variable is the level of intrapartum of fetal scalp pH, while in the second column is an indicator equal to one if the intrapartum pH is below 7.20. All specifications include maternal and pregnancy controls, and weekday and hospital fixed effects. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. The sample is restricted to single births, unscheduled c-sections, and vaginal deliveries (excluding breech vaginal babies) for which we have information about the intrapartum pH. Standard errors (in parentheses) are clustered at the hospital-shift level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A10: IV Estimation – Apgar Scores: Comparing C-Sections with Eutocic Births

	Apgar Score 1			Apgar Score 5		
	(1)	(2)	(3)	(1)	(2)	(3)
Panel A. 2SLS						
Unplanned CS	-1.179***	-1.218***	-1.161**	-0.907**	-0.954**	-0.942**
	(0.448)	(0.459)	(0.514)	(0.372)	(0.382)	(0.426)
Mean of Y		8.945			9.809	
Panel B. First stage						
Early night	0.090***	0.088***	0.078***	0.090***	0.088***	0.078***
	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)	(0.012)
Observations	4886	4886	4886	4884	4884	4884
First-stage F	45.329	43.974	39.192	45.222	43.852	39.102
Maternal controls		✓	✓		✓	✓
Pregnancy controls			✓			✓

Notes: The table shows the instrumental variables estimates of the effect of an unplanned cesarean birth on Apgar scores 1 and 5, respectively, compared to an eutocic birth (a vaginal birth without any instrumentation). The endogenous variable, an indicator for an unplanned cesarean birth, is instrumented with a dummy variable equal to one for births between 11 pm to 4 am (early night). Panel A shows the second stage coefficients, while Panel B displays the corresponding first stage results. First-stage F statistics are reported at the bottom of the table. The first column for each outcome shows the results of this regression controlling only for weekday and hospital fixed effects; in the second column maternal controls are added, and in the third column pregnancy controls are also included. Maternal controls comprise: level of education, nationality, maternal weight, height, age, and marital status. Pregnancy controls include: an indicator for previous c-section, the trimester in which prenatal care began, an indicator for obstetric risk, an indicator for preterm birth, and an indicator for induced labor. Mean of Y refers to the average of the outcome variable in the used sample. The sample is restricted to single births, unscheduled c-sections, and eutocic vaginal deliveries. Standard errors (in parentheses) are clustered at the hospital-shift level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$